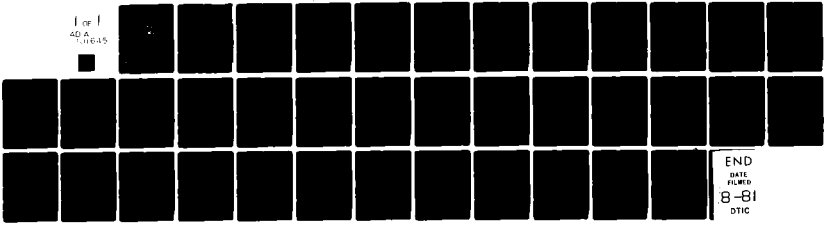
AD-A101 645 NAVAL POSTGRADUATE SCHOOL MONTEREY CA F/G 12/1
GENERATION OF SOME FIRST-ORDER AUTOREGRESSIVE MARKOVIAN SEQUENC--ETC(U)
MAR 81 A J LAWRANCE, P A LEWIS
UNCLASSIFIED NPS55-81-003 NL



NPS55-81-003

NAVAL POSTGRADUATE SCHOOL

Monterey, California

GENERATION OF SOME FIRST-ORDER AUTOREGRESSIVE MARKOVIAN SEQUENCES OF POSITIVE RANDOM VARIABLES WITH GIVEN MARGINAL DISTRIBUTIONS

by

A. J. Lawrance

P. A. W. Lewis

March 1981

Prepared for:
Naval Postgraduate School
Monterey, CA 93940

NAVAL POSTGRADUATE SCHOOL
MONTEREY, CALIFORNIA

Rear Admiral J. J. Ekelund
Superintendent

D. A. Schrady
Acting Provost

This report was prepared by:

A. J. Lawrance (P.A.W.L.)
A. J. Lawrance
University of Birmingham
Birmingham, England

P. A. W. Lewis
P. A. W. Lewis, Professor
Department of Operations Research

Reviewed by:

K. T. Marshall
K. T. Marshall, Chairman
Department of Operations Research

Released by:

William M. Tolles
William M. Tolles
Dean of Research

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE *(When Data Entered)*

REPORT DOCUMENTATION PAGE	READ INSTRUCTIONS BEFORE COMPLETING FORM
1. REPORT NUMBER: NPS55-81-003 (14)	2. GOVT ACCESSION NO.: AD-A107 645
4. TITLE *(and Subtitle)*: (6) GENERATION OF SOME FIRST-ORDER AUTOREGRESSIVE MARKOVIAN SEQUENCES OF POSITIVE RANDOM VARIABLES WITH GIVEN MARGINAL DISTRIBUTIONS	5. TYPE OF REPORT & PERIOD COVERED
	6. PERFORMING ORG. REPORT NUMBER
7. AUTHOR(s): (10) A. J. Lawrance P. A. W. Lewis	8. CONTRACT OR GRANT NUMBER(s)
9. PERFORMING ORGANIZATION NAME AND ADDRESS: Naval Postgraduate School, Monterey, CA 93940	10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS: (12) 32
11. CONTROLLING OFFICE NAME AND ADDRESS: Naval Postgraduate School, Monterey, CA 93940 — 11) May 81	12. REPORT DATE: March 1981
	13. NUMBER OF PAGES: 34
14. MONITORING AGENCY NAME & ADDRESS *(if different from Controlling Office)*	15. SECURITY CLASS. *(of this report)*: Unclassified
	15a. DECLASSIFICATION/DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT *(of this Report)*

Approved for public release; distribution unlimited.

17. DISTRIBUTION STATEMENT *(of the abstract entered in Block 20, if different from Report)*

18. SUPPLEMENTARY NOTES

19. KEY WORDS *(Continue on reverse side if necessary and identify by block number)*

Markovian Sequences; Autoregressive; Positive Random Variables; Exponential processes; NEAR(1) process; Gamma first-order autoregressive process; Gamma-Beta process; NMEAR(1) process.

20. ABSTRACT *(Continue on reverse side if necessary and identify by block number)*

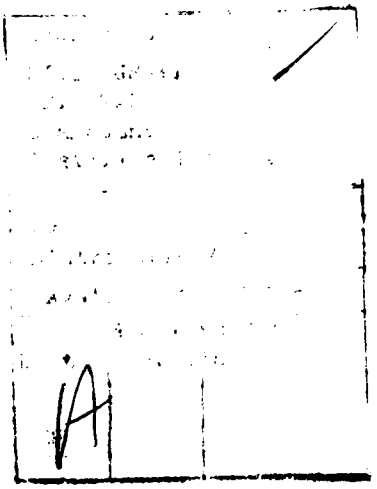
Methods for simulating dependent sequences of continuous positive-valued random variables with exponential uniform, Gamma, and mixed exponential marginal distributions are given. In most cases the sequences are first-order, linear autoregressive, Markovian processes. A very broad two-parameter family of this type, GNEAR(1), with exponential marginals and both positive and negative correlation is defined and its transformation to a similar multiplicative process with uniform marginals is given. It is shown that for a

DD FORM 1 JAN 73 1473 EDITION OF 1 NOV 65 IS OBSOLETE
S/N 0102-014-6601

SECURITY CLASSIFICATION OF THIS PAGE *(When Data Entered)*

Abstract Cont'd.

subclass of this two-parameter family extension to mixed exponential marginals is possible, giving a model of broad applicability for analyzing data and modelling stochastic systems, although negative correlation is more difficult to obtain than in the exponential case. Finally, two schemes for autoregressive sequences with Gamma distributed marginals are outlined. Efficient simulation of some of these schemes is discussed.



Generation of Some First-order Autoregressive Markovian Sequences of Positive Random Variables with given Marginal Distributions

by

A. J. Lawrance
Dept. Statistics, University of Birmingham, Birmingham, England

and

P. A. W. Lewis
Dept. Operations Research, Naval Postgraduate School, Monterey, CA

ABSTRACT

Methods for simulating dependent sequences of continuous positive-valued random variables with exponential uniform, Gamma, and mixed exponential marginal distributions are given. In most cases the sequences are first-order, linear autoregressive, Markovian processes. A very broad two-parameter family of this type, GNEAR(1), with exponential marginals and both positive and negative correlation is defined and its transformation to a similar multiplicative process with uniform marginals is given. It is shown that for a subclass of this two-parameter family extension to mixed exponential marginals is possible, giving a model of broad applicability for analyzing data and modelling stochastic systems, although negative correlation is more difficult to obtain than in the exponential case. Finally, two schemes for autoregressive sequences with Gamma distributed marginals are outlined. Efficient simulation of some of these schemes is discussed.

1. INTRODUCTION

In a recent series of papers [1,2,3,4,5,6,7,8,9] some simple models have been derived for stationary dependent sequences of positive, continuous random variables with given first-order marginal distributions. In general the dependency structure, as measured by second-order joint moments (serial correlations) mimics that of the usual linear mixed auto-regressive-moving average (ARMA) models which have been used for so long in time-series analysis. In the ARMA models, which are defined quite generally, there is in usage an implicit assumption of marginal normality of the random variables. This is clearly not the case if the random variables are positive, say the times between events in a series of events (Cox and Lewis [10]) or the successive response times at a computer terminal. Thus the new models are derived to accommodate situations in which the dependent random variables have, for instance, exponential, Gamma, uniform and mixed exponential marginal distributions. The exponential case is the most highly developed, with the nomenclature (Lawrance and Lewis [4]) EARMA (p,q) (exponential process with mixed moving average-autoregressive structures of orders p and q respectively) and NEARMA(p,q) (new EARMA(p,q)). A generalization to extend the range of attainable autocorrelations to negative values has been defined with the nomenclature GNEARMA(p,q).

The development of the probabilistic properties of these processes is given in the referenced papers, applications

to queueing models and computer systems modelling by Lewis and Shedler [11] and Jacobs [12,13], while development of estimation and testing procedures has just begun.

The object of the present paper is to define and discuss the simulation of the processes on digital computers, though for the sake of brevity only the first-order Markovian, autoregressive case is considered. The simplicity of structure of these models--in general they are linear additive mixtures of random variables--makes them ideal for this purpose. However, stationarity conditions are sometimes difficult to derive analytically and in some cases it is not simple to generate the innovation random variables in the processes. A striking example of this is the case of the Gamma first-order autoregressive process, given in Section 4A, for which an efficient means of simulation was reported by Lawrance [7] for some parametric values. This procedure carries over into another Gamma process, the Gamma-Beta process, which will be discussed in Section 4B.

In Section 3 it is shown that a simple transformation of the exponential sequences gives a direct multiplicative method for generating dependent processes with uniform marginals. These could be the basis in simulations for many other types of dependent sequences.

Finally the NMEAR(1) process is detailed in Section 4C; this generates a first-order Markovian process with mixed exponential marginals. It is useful for simulating situations in which the observed random variables are correlated and overdispersed relative to an exponentially distributed random variable.

2. EXPONENTIAL AUTOREGRESSIVE MARKOVIAN SEQUENCES

We give here three methods of generating first-order autoregressive, Markovian sequences with exponential marginal distributions. The first two are defective in terms of their sample path properties (the first more so than the second) while the third, NEAR(1) and its generalization GNEAR(1), is satisfactory in this respect and is a very rich model. The defect of the first two sequences is also highlighted by the simulation procedures used; they can be generated from one sequence of exponential variables.

The word "autoregression" in the context of a stochastic sequence $\{X_n\}$ is often used rather vaguely. In the first place <u>linear</u>, <u>additive</u> autoregression is usually implied. In the second place first-order autoregression can mean that in the defining equation for X_n the previous value enters explicitly. Thirdly, it can mean that the conditional expectation of X_n, given $X_{n-1} = x_{n-1}$, is an additive linear function of x_{n-1};

$$E(X_n | X_{n-1} = x_{n-1}) = a + bx_{n-1}. \tag{2.1}$$

The processes discussed in this paper are autoregressive in the latter two senses and, except in the case of uniform marginals, are autoregressive in a linear additive way. They are also Markovian; the Markovian property (first-order) means that the probability structure of $X_n, X_{n+1}, \ldots$, given $X_{n-1} = x_n$ is independent of $X_{n-2}, X_{n-3}, \ldots$.

2A. The Exponential DAR(1) Process

A very simple exponential autoregressive Markovian sequence is generated by the equation (Jacobs and Lewis [14,15])

$$X_n = V_n X_{n-1} + (1 - V_n)E_n \ , \tag{2.2}$$

where the V_n's, $n = 1,2,\ldots$ are i.i.d with $P\{V_n=1\} = 1 - P\{V_n=0\} = \rho$ and E_n, $n = 1,2,\ldots$ are, as throughout the paper, independent exponential random variables with parameter λ and independent of the V_n's; that is

$$\begin{aligned} P\{E_n \leq x\} &= 1 - e^{-\lambda x}, \quad x \geq 0, \quad \lambda > 0 \\ &= 0 \quad , \quad x < 0. \end{aligned} \tag{2.3}$$

For this process the serial correlations $\rho_k = \mathrm{corr}(X_n, X_{n+k})$ are

$$\rho_k = \rho^k \tag{2.4}$$

and

$$E(X_n | X_{n-1} = x_{n-1}) = \rho_1 x_{n-1} + (1 - \rho_1)/\lambda. \tag{2.5}$$

This process, which was introduced to model discrete valued variables, is not well suited to modelling continuous data because runs of X_n's with the same value can occur quite frequently in the sample paths of the process. This happens

when X_{n-1} is picked successively in (2.2), rather than the innovation E_n. Moreover the lengths of the runs of similar values are geometrically distributed.

2B. The Exponential EAR(1) Process

Another model is derived from the usual linear model

$$X_n = \rho X_{n-1} + \varepsilon_n \tag{2.6}$$

in which the i.i.d. innovation process $\{\varepsilon_n\}$ is chosen so that the X_n's are marginally exponential(λ). Gaver and Lewis [1] show that for this to be true, one must have $0 \leq \rho < 1$ and

$$\varepsilon_n = \begin{cases} E_n & \text{w.p. } 1-\rho, \\ 0 & \text{w.p. } \rho\ , \end{cases} \tag{2.7}$$

where $\{E_n\}$, as previously, are i.i.d. exponential(λ). Again $\rho_k = \rho^k$ and $E(X_n | X_{n-1} = x_{n-1}) = \rho_1 x_{n-1} + (1-\rho_1)\lambda$, as at (2.4) and (2.5) for the exponential DAR(1) model. The difference is in the sample paths; the EAR(1) process simulations show runs of geometrically decreasing X_n's, but no runs of constant value. The geometrically distributed runs occur when only ρX_{n-1} is picked in (2.6).

The Markov property of the two sequences implies that if X_0 is chosen to be E_0, an exponential(λ) random variable independent of $E_1, E_2, \ldots$, then $X_1, X_2, \ldots$ forms a stationary sequence.

Naive inspection of the defining equations (2.2), (2.6) and (2.7) suggest that to generate a stationary sequence of length N, $X_1, \ldots, X_N$, $(N+1)$ i.i.d. exponential deviates and N uniform variates (for the selection process) are needed. However, the sequences can be generated from only one exponential sequence; this is possibly related to the degeneracy in the processes. This method uses the memoryless property of exponential(λ) variables, namely that if E_n is given to be greater than a constant γ, then $E_n - \gamma$ is again exponential(λ).

Thus the algorithm for the EAR(1) process is to initialize by setting $X_0 = E_0$; subsequently set $X_n = \rho X_{n-1}$ if $E_n \leq x_\rho = -\ell n(1-\rho)/\lambda$; otherwise set $X_n = \rho X_{n-1} + (E_n - x_\rho)$. This uses the fact that, from (2.3), $P\{E_n \leq x_\rho\} = \rho$.

Even greater efficiency can be obtained, though this must be qualified by considerations as to whether (i) the X_n's are to be generated one at a time or in an array; (ii) a subroutine is available to generate exponential random variables faster than can be done by taking logarithms of uniform deviates, and (iii) the speed of division in the computer is short compared to the time needed for generation of uniform deviates.

The more efficient scheme recycles uniform variables, i.e. if U is given to be between constants a and b, where $0 < a < b \leq 1$, then $(U-a)/(b-a)$ is a uniform random variable. (Note that its value is not given, only that it is in (a,b)). The expected number of uniform deviates required

to generate an EAR(1) process of length N with this algorithm is $1 + (1-\rho)N$, which is less than the number N required to generate an i.i.d. exponential(λ) sequence. Also the expected number of logarithms is $(1-\rho)N$, while N comparisons are always needed.

2C. The Exponential NEAR(1) Process

A broader two-parameter exponential sequence which is a first-order autoregressive, Markovian process and an additive linear mixture of random variables is given by Lawrance [7] and developed by Lawrance and Lewis [5]. Called NEAR(1), the sequence is defined as

$$X_n = \varepsilon_n + \begin{cases} \beta X_{n-1} & \text{w.p.} \quad \alpha \\ 0 & \text{w.p.} \quad 1-\alpha \end{cases} \qquad n = 1,2,\ldots, \tag{2.8}$$

where $0 \leq \alpha \leq 1$ and $0 \leq \beta \leq 1$ but $\alpha = \beta \neq 1$. Also the selection process is done independently for each n. It can be shown that for the X_n to be marginally exponential(λ) the innovation variable ε_n must be generated from an E_n by the exponential mixture

$$\varepsilon_n = \begin{cases} E_n & \text{w.p.} \quad \delta = \dfrac{1-\beta}{1-(1-\alpha)\beta} \\ (1-\alpha)\beta E_n & \text{w.p.} \quad 1-\delta = \dfrac{\alpha\beta}{1-(1-\alpha)\beta} \end{cases} \qquad n = 1,2,\ldots \tag{2.9}$$

providing α and β are not both equal to one. When $\alpha = 0$ or $\beta = 0$ the $\{X_n\}$ are i.i.d. exponential variables, whereas when $\alpha = 1$ the EAR(1) model given at (2.6) and (2.7) is obtained. In fact choosing α as a function of β in a suitable way, e.g. $\beta = \alpha$, gives an exponential model with a full positive range of serial correlation of order one, since it is easily shown that

$$\rho_k = (\alpha\beta)^k . \tag{2.10}$$

Again

$$E(X_n | X_{n-1} = x_{n-1}) = \alpha\beta x_{n-1} + (1-\alpha\beta)/\lambda$$

$$= \rho_1 x_{n-1} + (1-\rho_1)/\lambda \tag{2.11}$$

and $X_0 = E_0$ gives a stationary sequence. Thus the correlations and regressions are the same as for the first two models. However the NEAR(1) process allows one to model a broader class of exponential sequences, as measured either by sample path behavior or higher-order joint moments; see Lawrance and Lewis [5] for details.

A particularly simple case occurs when $\beta = 1$; this model, called TEAR(1), is very tractable analytically and, as will be shown in Section 4C, extends easily to the case of mixed exponential distributions for the X_n.

Note that in the NEAR(1) process the innovation ε_n is always present unless $\alpha = 1$ and it is therefore not

possible to simulate the stationary process with less than N+1 uniform variates. A detailed algorithm is given in Section 2E for a more general case. Since for a stationary array of N X_n's, exactly N+1 uniform deviates are required because of the ability to recycle the uniforms and transform them into exponentials, it could be advantageous to generate these uniform deviates in an array which would be replaced one at a time by the X_n's. Care must be taken with the recycling of the uniform variates U if $\gamma = 1-\alpha$ is close to one or zero. In that case it is probably better for computational reasons to use 2(N+1) uniform variates. Note that $\alpha = 0$ gives the EAR(1) process. When $\beta = 1$ a simpler algorithm can be used since E_n is no longer a mixture of two exponentials; this important special case is called the TEAR(1) exponential process.

When $\beta = \frac{1}{2-\alpha}$ another one-parameter sub-class of the NEAR(1) process is obtained with strikingly regular sample path properties. For this process $P\{X_n > X_{n-1}\} = 1/2$; this is in striking contrast to the EAR(1) process whose sample paths show runs-down and the TEAR(1) process whose sample paths show runs-up. This is illustrated in Figure 1; all sequences have $\rho = 0.75$ and are transformations of the same E_n sequence. The parameter space of the NEAR(1) process is illustrated in Figure 2.

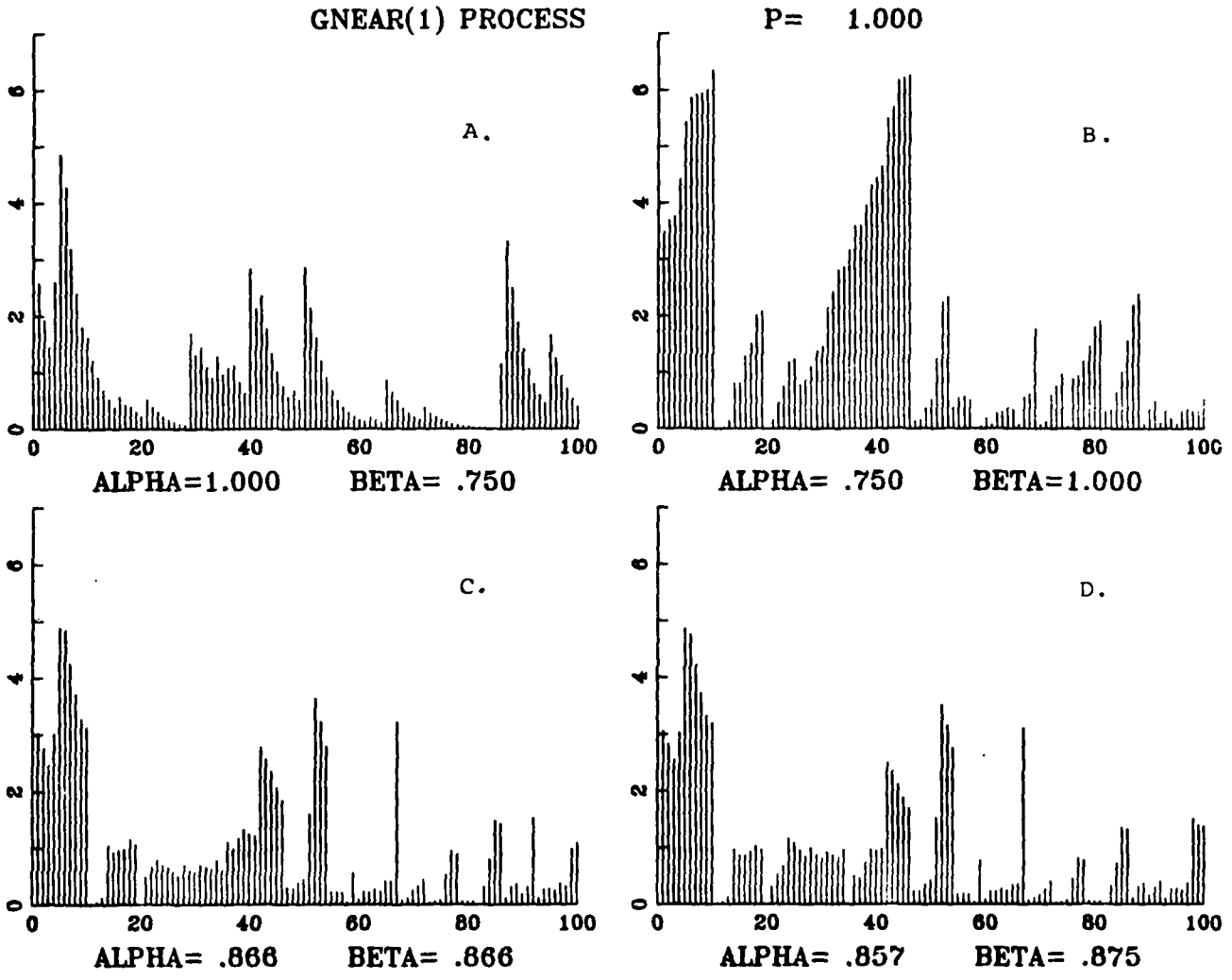


Figure 1. Sample paths of NEAR(1) processes, all with $\alpha\beta = \rho_1 = 0.75$, for different values of α and β. Figure A is the EAR(1) process ($\alpha = 1$, $\beta = 0.75$), Figure B. is the TEAR(1) process ($\alpha = 0.75$, $\beta = 1$), Figure C is the PREAR(1) process $\alpha = \frac{1}{2-\beta} = 0.857$, and Figure D is the REAR(1) process $\alpha = \beta = (0.75)^{1/2}$. All the sample paths are transformations of the same i.i.d. exponential sequence E_n, $n = 0,1,\ldots,100$.

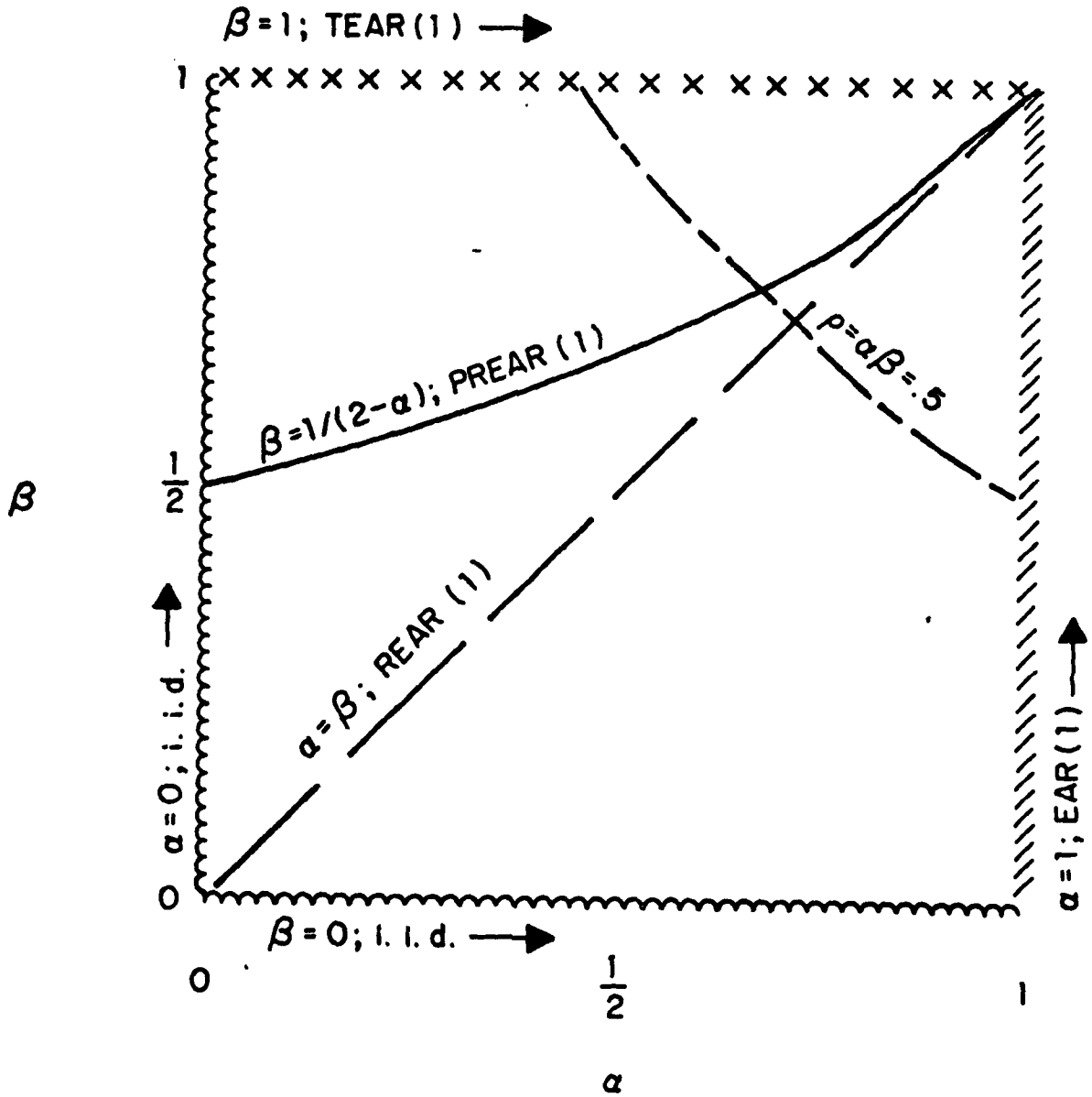


Figure 2. Parameter space for the NEAR(1) exponential autoregressive process. The cases $\alpha = 0$ and/or $\beta = 0$ give i.i.d. exponential sequences while $\beta = 1/(2-\alpha)$ gives the partially reversible PREAR(1) process for which $P(X_n < X_{n-1}) = 1/2$. For $\alpha = 1$ we have the original exponential process EAR(1) which tends to have runs-down. The TEAR(1) case, $\beta = 1$ gives very simple analytics but exhibits runs up. Also shown is a locus of constant ρ_1, in this case $\rho_1 = \alpha\beta = .5$.

2D. The Generalized Exponential Process GNEAR(1) with Negative Correlation

The exponential processes defined in Sections 2A, 2B, and 2C do not exhibit negative correlation or alternation of correlations. Such behaviour is found in, say, a normal linear first-order process for which $\rho_j = \text{cor}(X_n, X_{n+j}) = \rho^j$ and $-1 < \rho < 1$ so that, for instance, ρ_1 can be negative. A scheme for broadening the correlation structure of the EAR(1) process is given in Gaver and Lewis [1]. However in the exponential case a much simpler alternative method is available.

Assume for simplicity that $\lambda = 1$. Now X_{n-1} is a unit exponential variable and $U_n = F(X_{n-1}) = 1 - \exp(-X_{n-1})$ is a uniform (0,1) variable, as is $1 - U_n = \exp(-X_{n-1})$. Then $X'_{n-1} = F^{-1}(1 - U_n) = F^{-1}(1 - F(X_{n-1})) = -\ell n(1 - \exp(-X_{n-1}))$ is a unit exponential variable; in fact it is the antithetic of X_{n-1} which gives the maximum negative correlation attainable in a bivariate exponential distribution:

$$r = \text{corr}(X_{n-1}, X'_{n-1}) = 1 - \pi^2/6 = -.6449. \tag{2.12}$$

Now the process

$$X_n = \begin{cases} \varepsilon_n + \beta X'_{n-1} = \varepsilon_n - \beta\, \ell n(1 - e^{-X_{n-1}}) & \text{w.p. } \alpha \\ \varepsilon_n & \text{w.p. } 1-\alpha \end{cases} \tag{2.13}$$

or

$$X_n = \varepsilon_n + U_n X'_{n-1} \tag{2.14}$$

in which the U_n's are i.i.d. with $P\{U_n=1\} = 1 - P\{U_n = 0\} = \alpha$ gives a process with autocorrelations which alternate in sign. In particular $\rho_1 = r(\alpha\beta)$. To combine this with the positive correlation case in a continuous way we introduce a new parameter $p \in [0,1]$ and i.i.d. indicator variables I_n, independent of U_n, from which $P\{I_n = 1\} = 1 - P\{I_n = 0\} = p$. Then the GNEAR(1) model is defined as

$$X_n = \varepsilon_n + U_n\beta\{I_nX_{n-1} + (1 - I_n)[- \ell n(1 - e^{-X_{n-1}})]\} \qquad (2.15)$$

and gives a complete range of first-order serial correlations

$$\rho_1 = \alpha\beta[p + (1-p)r]. \qquad (2.16)$$

Higher lag correlations are more complicated and will be given elsewhere. Note, however, that $1-p = 1/(1-r)$ gives a case in which X_n and X_{n-1} are a bivariate exponential pair which are dependent but have zero correlation.

Figure 3 give four sample paths for the case $p = 0$ for values of α and β corresponding to those in Figure 1, which is the case $p = 1$.

2E. Algorithms for the GNEAR(1) Process

We give here two algorithms for the GNEAR(1) process, one based on generation of uniform deviates one at a time, the other based on the availability of subroutines to generate arrays of uniform and exponential variables. The sequences generated are unit exponentials ($\lambda = 1$). The special case $\alpha = 1$ (EAR(1)) and $\beta = 1$ (TEAR(1)) are handled separately as they will cause divides by zero in the main algorithm.

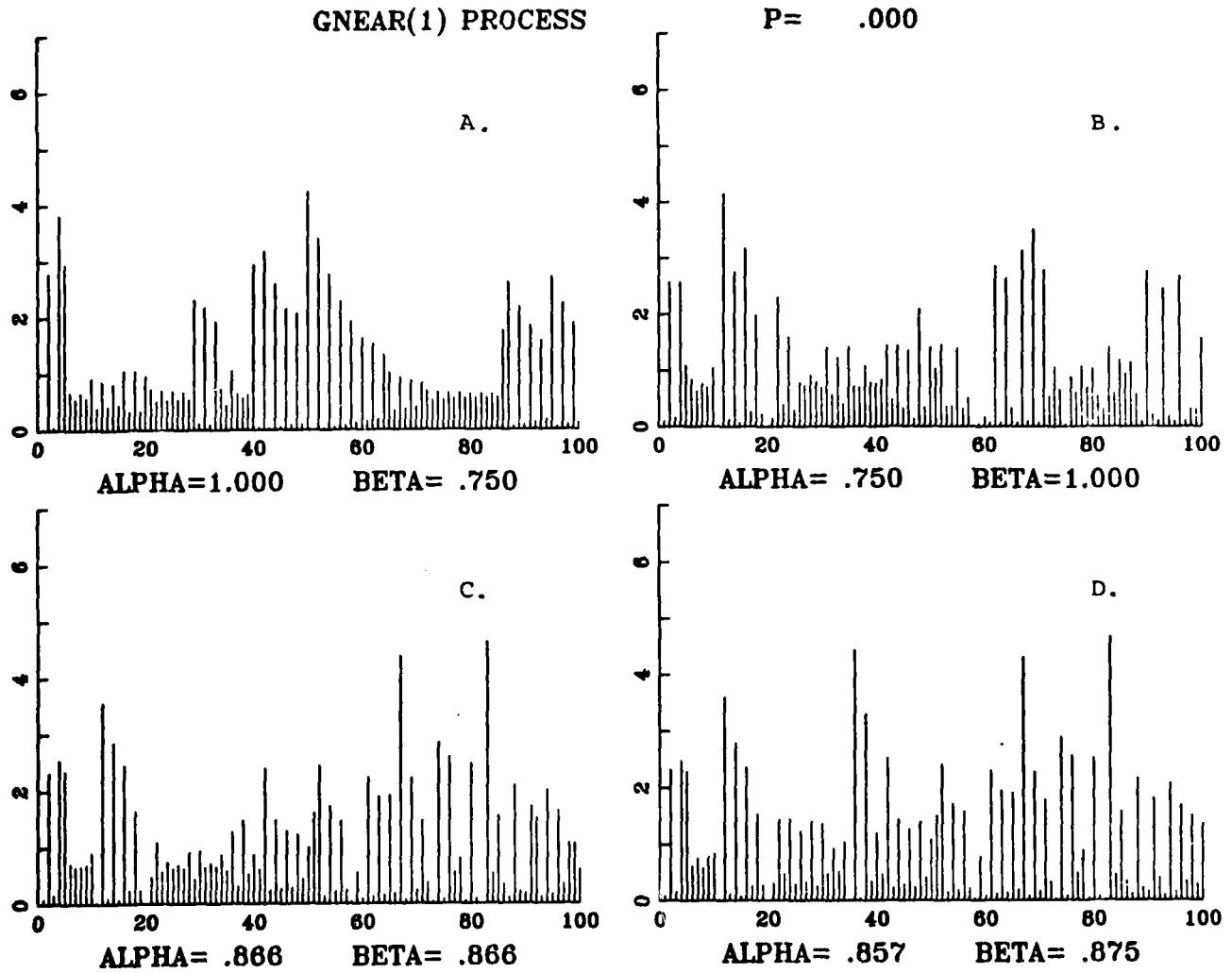


Figure 3. Sample paths for the GNEAR(1) process with $p = 0$, corresponding to those in Figure 1 for the NEAR(1) process ($p = 1$). Here $\rho_1 = -0.6449\ \alpha\beta$.

ALGORITHM GNEAR1A

This algorithm generates a sample of size N from the GNEAR1 process. It is based on the generation of uniform random numbers one at the time, with recycling of these uniform random numbers for further use. It is assumed that a subroutine (UNIFORM) exists that generates raw uniforms. Input values are N, ALPHA, BETA and P the last three taking values in the closed interval [0,1]. However ALPHA = BETA = 1 is not allowed.

```
INPUT  N, ALPHA, BETA, P
CALL  UNIFORM (U)                          /* Generate uniform U*/
X(0) ← -log_e(U)                           /* convert to exponential*/
C ← (1-ALPHA)*BETA
IF  ALPHA = 0          THEN     D ← 1
                       ELSE     D ← (1-BETA)/(1-C)
END IF
DO   I ← 1  to  N
     CALL UNIFORM (U)
     IF  U ≤ ALPHA     THEN     CALL UNIFORM (V)
                                IF V ≤ P THEN Y ← BETA* X(I-1)
                                         ELSE Y ← - BETA* log_e(1-exp(-X(I-1)))
                                END IF
                                U ← U/ALPHA                 /* Recycle U */
                       ELSE     Y ← 0
                                U ← (U-ALPHA)/(1-ALPHA)     /* Recycle U */
     END IF
     IF  U ≤ D         THEN     Z ← -log_e(U/D)
                       ELSE     Z ← - C * log_e(U-D)/(1-D))
     END IF
     X(I) ← Z + Y
END DO
```

ALGORITHM GNEAR1B

This algorithm generates a sample of size N from the CNEAR1 process. It is based on the generation of arrays of uniform and exponentially distributed random numbers. It is assumed that subroutines UNIFORM and EXPON exist that generate arrays of uniform random numbers and exponential random numbers respectively. For each GNEAR1 number generated two raw uniform random numbers are used. The first is recycled in those cases where a third uniform is needed to make the selection with probability P. Input values are N, ALPHA, BETA and P the last three taking values in the closed interval [0,1], except the case ALPHA = BETA = 1

```
INPUT  N, ALPHA, BETA, P
C ← (1 - ALPHA)*BETA
IF ALPHA = 0          THEN    D ← 1
                      ELSE    D ← (1-BETA)/(1-C)
CALL UNIFORM (U, 2*N)         /*Generate array of 2N uniforms*/
CALL EXPON (E, N+1)           /*Generate array of N+1 exponentials*/
X(0) ← E(N+1)
DO   I ← 1  to  N
     IF  U(I) ≤ ALPHA  THEN   V = U(I)/ALPHA              /* Recyle U */
                              IF V ≤ P THEN Y ← BETA* X(I-1)
                                       ELSE Y ← - BETA* log_e(1-exp(-X(I-1)))
                              END IF
                      ELSE    Y ← 0
     END IF
     IF  U(I+I) ≤ D   THEN    Z ← E(I)
                      ELSE    Z ← C*E(I)
     END IF
     X(I) ← Z + Y
END DO
```

3. UNIFORM MARKOVIAN SEQUENCES, NUAR(1)

It is convenient to have dependent sequences of random variables with marginal distributions other than exponential. Before discussing other solutions to Equation (2.8) we show that a simple transformation of the NEAR(1) process gives a two-parameter family of Markovian random variables with uniform marginal distributions. It is well-known that an exponential transformation of a unit exponential random variable gives a uniformly distributed random variable. Thus we have from (2.8) and (2.9) the multiplicative model for a uniform Markovian sequence $\{X_n\}$, $n = 1,2,\ldots$, called NUAR(1),

$$X_n = \begin{cases} \varepsilon_n \, X_{n-1}^{\beta} & \text{w.p. } \alpha \\ \varepsilon_n & \text{w.p. } (1-\alpha) \end{cases} \qquad (3.1)$$

$$n = 1,2,\ldots,$$

where

$$\varepsilon_n = \begin{cases} U_n & \text{w.p. } \gamma = \dfrac{1-\beta}{1-(1-\alpha)\beta} \qquad (3.2) \\ U_n^{(1-\alpha)\beta} & \text{w.p. } 1-\gamma = \dfrac{\alpha\beta}{1-(1-\alpha)\beta} \qquad (3.3) \end{cases}$$

$$n = 1,2,\ldots$$

for U_n, $n = 1,2,\ldots$, i.i.d. uniformly distributed, providing that α and β are not both equal to one. Again if X_0 is uniformly distributed and independent of $U_1, U_2, \ldots$ the sequence is stationary.

The sequence is clearly quite simply extended to give negative correlation, as in the GNEAR(1) process; in fact X_{n-1} in (3.1) is just replaced by $(1-X_{n-1})$. Algorithms are easily obtained by adaptation of those given in Section 2E for the exponential case. It remains to find the correlation structure and the regression of X_n on X_{n-1}.

To do the former, let X_n^* be a NEAR(1) sequence with $\lambda = 1$, so that the sequence X_n at (3.1) is given by $X_n = \exp\{-X_n^*\}$. Now the joint Laplace-Stieltjes transform of X_n^* and X_{n-k}^*, $\phi_{X_n^*,\, X_{n-k}^*}(s,t) = E\{\exp[-sX_n^* - tX_{n-k}^*]\}$, is given by Lawrance and Lewis [5]. Setting $s = t = 1$ in $\phi_{X_n^*,\, X_{n-k}^*}(s,t)$ gives,

$$\phi_{X_n^*,\, X_{n-k}^*}(1,1) = E\{\exp(-X_n^*)\ \exp(-X_{n-1}^*)\}$$

$$= E(X_n X_{n-k}) \ . \tag{3.4}$$

Then using the fact that for a uniform random variable $E(X) = 1/2$ and $\text{var}(X) = 1/12$ we have, after simplification,

$$\rho_k = \text{corr}(X_n, X_{n-k}) = \frac{3}{2 + \beta^k} \sum_{i=1}^{k} \left(\frac{\alpha\beta}{1 + (1-\alpha)\beta^i} \right)$$

$$k = 1,2,\ldots \ . \tag{3.5}$$

Note that this is not simply a geometrically decaying correlation sequence, as for the NEAR(1) process. However, for the important special case when $\beta = 1$ we get

$$\rho_k = \left(\frac{\alpha}{2-\alpha}\right)^k , \qquad k = 1,2,\ldots , \tag{3.6}$$

and thus the serial correlations ρ_k are the kth power of ρ_1, which takes on any value between 0 and 1. Thus we have a particularly simple uniform Markovian sequence, although the sample paths will tend to have runs-up.

A similar analysis given in Lawrance and Lewis [5] shows that

$$E(X_n | X_{n-1} = u) = \frac{1}{2} \frac{1 + \beta}{\{(1 + (1-\alpha)\beta\}} \{1 - \alpha + \alpha u^{\beta}\} \tag{3.7}$$

so that the regression is not linear for this Markov process with uniform marginals, unless $\beta = 1$.

This uniform sequence could form the basis, via a probability integral transform, of many other sequences with given marginals. The parametrization $\beta = \frac{1}{2-\alpha}$ is a good choice for a one parameter model since this case gives a sample path which is partially time-reversible (see Lawrance and Lewis [5]), with a balance of runs-up and runs-down. However, marginal transformations do not preserve correlation structure, as shown at (3.5), and it is therefore useful to see whether sequences with marginals other than exponential can be generated from (2.8); this requires finding, if possible, a suitable choice of innovation sequence ε_n. The result will be a simple process with autoregressive Markovian structure and the desired marginal distribution.

4. MARKOVIAN SEQUENCES WITH SOME OTHER MARGINALS

Although an exponential distribution is a common assumption for positive random variables met with in problems in operations research, it is too narrow an assumption to encompass many real situations. Therefore parametric distribution models are invoked which include the exponential as a special case and which allow for the modelling of data which has greater or lesser dispersion than exponentially distributed data. Two commonly used models are

(i) the Gamma(k,λ) distribution whose probability density function is

$$f(x) = \frac{\lambda(\lambda x)^{k-1}e^{-\lambda x}}{\Gamma(k)}, \quad k > 0;\ \lambda > 0;\ x \geq 0, \tag{4.1}$$

where $\Gamma(k)$ is the complete gamma function, and

(ii) the (convex) mixture of exponential random variables

$$f(x) = \pi_1\lambda_1 e^{-\lambda_1 x} + (1-\pi_1)e^{-\lambda_2 x}, \quad 0 < \lambda_1 < \lambda_2;$$

$$x \geq 0,\ 0 \leq \pi_1 \leq 1 . \tag{4.2}$$

The Gamma distribution has dispersion, measured by the coefficient of variation $C(X) = \sigma(X/E(X))$, which is greater than the exponential value of 1 if $k < 0$ and less than 1 if $k > 1$. The mixed exponential always has $C(X) \geq 1$, the equality occurring when the special case of an exponential random variable with parameters λ_1 or λ_2 holds.

4A. The Gamma GAR(1) process

The solution of the standard first-order autoregressive equation (2.6) with stationary gamma marginals defines the GAR(1) process. Using Laplace-Stieltjes transforms with (2.6) shows that for X_n to be Gamma (k,λ), we must have

$$\phi_\varepsilon(s) = E(e^{-s\varepsilon}) = \left\{\rho + (1-\rho)\ \frac{\lambda}{\lambda+s}\right\}^k . \tag{4.3}$$

For k integer this has an explicit inverse. For example, for $k = 2$ the innovation ε is zero with probability ρ^2, is exponential(λ) with probability $2\rho(1-\rho)$ and is Gamma$(2,\lambda)$ with probability $(1-\rho)^2$. It is easy to show in general that ε is zero with probability ρ^k, so that the "zero defect" is not serious for large k. A method of simulating a random variable whose Laplace-Stieltjes transform is equation (4.3) was derived by Lawrance [7], using the fact that this sequence arises in a particular type of shot noise process. From this we have the

Gamma Innovation Theorem

Let N be a Poisson random variable with parameter $\theta = -k\ \ell n(\rho)$. Let $U_1, U_2, \ldots, U_N$ be uniformly distributed over $(0,1)$ and independent. Let $Y_1, \ldots, Y_N$ be exponential(λ) and independent. Then ε can be simulated using

$$\varepsilon \begin{cases} = \sum_{m=1}^{N} Y_m \rho^{U_m} & \text{if } N > 0 , \\ = 0 & \text{if } N = 0 . \end{cases} \tag{4.4}$$

A proof is not given here. Note that ε is zero with probability $\exp\{-k \ln(\rho)\} = \rho^k$. Also the Poisson number N of uniform and exponential random variables which must be generated for each ε has expected value $\theta = -k \ln(\rho)$. This will be prohibitively large, and the simulation will be very inefficient, if k is large and/or ρ is close to zero. Neither of these cases is serious, however. If k is large, say greater than 50, the sequence is almost normal and the usual normally distributed, AR(1) linear process can be used. If ρ is as small as 0.001 then $E(N)$ is only $k \times (6.9078)$ which is still reasonable. However, for ρ this small the sequence is approximately i.i.d. Gamma and acceptance-rejection techniques for simulating Gamma variables are known.

It is quite simple to write algorithms for the GAR(1) case analogous to those in Section 2E. It would pay to have a built-in routine for generating the Poisson variable which will bypass further calculations if $N = 0$. In other words routines for generating Poisson variates which start by searching at the median of a table of cumulative Poisson probabilities will be inefficient.

Unfortunately the NEAR(1) process does not appear to extend to the Gamma case; it can be shown explicitly that there is no innovation ε_n in equation (2.8) which will make X_n have a Gamma distribution with $k = 2$ if $\alpha \neq 1$.

There is, however, another model, the Gamma-Beta model, inspired by an example in Verwaart [16] and discovered independently by Fishman [17], which is quite broad and which can be simulated using the Gamma Innovation Theorem.

4B. The Gamma-Beta Model, GBAR(1)

This model is a linear autoregressive process with random coefficients which includes the GAR(1) process but is of limited practical use in data analysis because its likelihood function is analytically untractable. Nevertheless it could be useful in simulations, particularly if the zero-defect in the GAR(1) process is unacceptable.

Thus, we define the stationary sequence as

$$X_n = \beta B_n X_{n-1} + \varepsilon_n \qquad 0 \leq \beta \leq 1 \ , \quad n = 0, \pm 1, \pm 2, \ldots, \quad (4.5)$$

where the B_n's are i.i.d. and independent of X_{n-1}, and B_n, for X_n to have a Gamma (k,λ) distribution, has a Beta(k-b,b) distribution with mean $E(B_n) = (k-b)/k$. The density is

$$f_{B_n}(x) = \frac{\Gamma(k)}{\Gamma(k-b)\Gamma(b)} x^{k-b-1} (1-x)^{b-1}. \qquad \begin{matrix} 0 \leq x \leq 1 \ . \\ 0 < b \leq k \ . \end{matrix} \qquad (4.6)$$

The distribution of the i.i.d. ε_n's to make X_n Gamma (k,λ) is still to be determined. To do this and to see the rationale behind the model, it is simplest to obtain the distribution of $B_n X_{n-1}$.

Now B_n can be generated as $Z_1/(Z_1+Z_2)$ where Z_1 and Z_2 are independent and Gamma $(k-b,\lambda)$ and Gamma (b,λ) respectively. Moreover, B_n is independent of (Z_1+Z_2), which could be used to generate X_{n-1}. Then, $B_n X_{n-1} = \{Z_1/(Z_1+Z_2)\}\{Z_1+Z_2\} = Z_1$ is Gamma $(k-b,\lambda)$. Using

this fact, which can also be shown analytically, and the defining equation (4.3) we have

$$E\left(e^{-sX_n}\right) = \phi_{X_n}(s) = \phi_{B_n X_{n-1}}(\beta s)\,\phi_{\varepsilon_n}(s) \quad , \tag{4.7}$$

so that on using the fact that for a Gamma(k,λ) variable the Laplace-Steiltjes transform is $(\lambda/\lambda+s)^k$, we have

$$\phi_{\varepsilon_n}(s) = \phi_{X_n}(s)/\phi_{B_n X_{n-1}}(\beta s)$$

$$= \left(\frac{\lambda}{\lambda+s}\right)^k \left(\frac{\lambda+\beta s}{\lambda+s}\right]^{k-b} \tag{4.8}$$

$$= \left(\frac{\lambda}{\lambda+s}\right)^b \left[\beta + (1-\beta)\,\frac{\lambda}{\lambda+s}\right]^{k-b} . \tag{4.9}$$

Thus ε_n is generated as the sum of a Gamma (b,λ) variate and, from (4.3), a Gamma innovation variable which can be generated from (4.4).

For this model

$$\rho_j = \text{corr}(X_n X_{n+j}) = \left(\beta\,\frac{k-b}{k}\right)^j \qquad j = 0,1,2,\ldots \tag{4.10}$$

and

$$E(X_n \; X_{n-1} = x_{n-1}) = \rho_1 x_{n-1} + (1-\rho_1)E(X_n).$$

Also when $b = 0$ we have the GAR(1) model. If $\beta = 1$ the model requires only a Gamma (b,λ) variate for ε_n and a

Beta(k-b,b) variate for B_n in the simulation. The case $k = 1$ gives an exponential process which is not the same as the NEAR(1) process.

It should be noted that while the Gamma Innovation Theorem makes the Gamma-Beta model tractable from a simulation viewpoint, it is still difficult to unite down a likelihood function or to get negative correlation. Thus further developments are needed from the Gamma case. An algorithm is given for the Gamma Beta Model on the next page.

4C. Mixed Exponential Markovian Processes MEAR(1) and TMEAR(1)

In addition to Gamma processes, first-order autoregressive Markovian processes with mixed exponential marginal distributions can be obtained from equations (2.8) and (2.9) in two special cases, and these sequences should be widely useful in modelling stochastic systems.

(i) The case $\alpha = 1$; MEAR(1).

In Gaver and Lewis [1] it is shown that the solution to the Laplace transform of ε_n for the linear model (2.6) is a constant ρ plus a (generally) non-convex mixture of three exponential functions. This can be shown to be a proper density function if $\rho \leq \lambda_1/\lambda_2$, but it can also be shown that it is not a density function for all ρ less than one and greater than or equal to zero. More particularly, Lawrance [6] showed that unless λ_1 is much smaller than λ_2 (and thus the X_n are very over-dispersed relative to an exponential random variable) a solution exists for ε_n for all ρ. Thus

ALGORITHM GBAR1 (GAMMA-BETA MODEL)

This algorithm generates a sample of size N from the GBAR1 process using array generators POISSON, UNIFORM, EXPONENTIAL and GAMMA to produce Poisson, Real Uniform (0,1), Exponential and Gamma distributed random numbers.

The following restrictions exist on the input parameters:

$$0 \leq B \leq K$$

$$0 < K$$

$$0 < BETA \leq 1$$

```
INPUT  N, BETA, B, K
TH ← -(K-B) * log_e(BETA)              /* Initialization */
CALL  POISSON(PSN,N,TH)    /*Generate  N  poisson deviates with parameter TH */
CALL  GAMMA(Z1,N,K-B)      /*Generate  N  Gammas with parameter K-B */
CALL  GAMMA(Z2,N,B)        /*Generate  N  Gammas with parameter B */
CALL  GAMMA(X(0), 1,K)     /* Initialize  X(0)  */
DO   I ← 1  to  N
     CALL  UNIFORM(U,PSN(I)           /* Generate PSN(I) real (0,1) uniforms */
     CALL  EXPON(E,PSN(I))            /* Generate PSN(I) Unit Exponentials */
     Y ← 0
     DO  J ← 1  TO  PSN(I)
         Y ← Y + E(J) * BETA ** U(J)
     END DO
     BN ← Z1(I)/(Z1(I) + Z2(I))       /* Compute BETA Deviate */
     X(I) ← BETA * BN * X(I-1) + Y
END DO
```

we have a useful process, although again the zero-defect of order ρ is a problem. However, one case which cannot be simulated this way occurs when $1/\lambda_2 = 0$, i.e. X is zero with probability $1-\pi_1$, and exponential (λ_1) otherwise. This kind of situation occurs in practice as, e.g., the waiting time for an item in an inventory system. Fortunately it can be handled in the next case.

(ii) The case $\beta = 1$; TMEAR(1).

When $\beta = 1$ in equation (2.8), a mixed exponential process TMEAR(1) is obtained which is extremely simple to simulate since the innovation ε_n is just the mixture of two exponentials for all $0 \leq \rho < 1$. Moreover, the process has no zero-defect. As discussed above, the sample paths will tend to "run up," but this is no great problem unless ρ is fairly large. Thus we have the following Theorem (Lawrance and Lewis [18]) which we state without proof:

TMEAR(1) Theorem

Let the first-order autoregressive, Markovian sequence $\{X_n\}$ be defined by

$$X_n = \varepsilon_n + V_n X_{n-1}, \qquad n = 1,2,3,...$$

where for the i.i.d. sequence V_n, $n = 1,2,3,...$, $P\{V_n=1\} = 1 - P\{V_n=0\} = \alpha$ for $0 \leq \alpha < 1$. Then the sequence $\{X_n\}$ is stationary and has a (convex) mixed exponential marginal distribution with probability density function

$$f_X(x) = \pi_1\lambda_1 e^{-\lambda_1 x} + \pi_2 e^{-\lambda_2 x} , \qquad x \geq 0 \tag{4.11}$$

where $0 < \lambda_1 < \lambda_2$; $0 < \pi_1 < 1$; $\pi_1 + \pi_2 = 1$ $\mu_1 = \frac{1}{\lambda_1}$ and $\mu_2 = \frac{1}{\lambda_2}$,

if ε_n is i.i.d. and has a (convex) mixed exponential distribution given by

$$f_\varepsilon(x) = \eta_1\gamma_1 e^{-\gamma_1 x} + \eta_2\gamma_2 e^{-\gamma_2 x} \qquad x \geq 0$$

with $\gamma_1 > \gamma_2 > 0$; $\eta_1, \eta_2 > 0$, $\eta_1 + \eta_2 = 1$ (4.12)

where

$$\mu = E(X) = \pi_2 \mu_2 + \pi_1\mu_1 ; \tag{4.13}$$

$$b = \mu_1 + \mu_2 - \alpha\mu ; \tag{4.14}$$

$$\beta = \mu_1 + \mu_2 - \mu ; \tag{4.15}$$

$$a = (1-\alpha)\mu_1\mu_2 ; \tag{4.16}$$

$$\gamma_1, \gamma_2 = \tfrac{1}{2}\{b \pm \sqrt{[b^2-4a]}\} ; \tag{4.17}$$

$$\gamma_0 = \pi_2\mu_1 + \pi_1\mu_2 ; \tag{4.18}$$

$$\eta_1 = (\gamma_1-\gamma_0)/(\gamma_1-\gamma_2) \; ; \; \eta_2 = (\gamma_2-\gamma_0)/(\gamma_2-\gamma_1) \tag{4.19}$$

and X_0 is independent of ε_1, ε_2, ... and has probability density function (4.11).

Note that the special cases where $\pi_1 = 0$ or $\pi_2 = 1$ give NEAR(1) exponential processes with parameters λ_2 and

λ_1 respectively. Thus they should be handled by Algorithm 2 since they will cause computational problems. The case $\lambda_1 = \lambda_2$ also gives a NEAR(1) process and is excluded for similar reasons. This guarantees that $\gamma_1 > \gamma_2$. The algorithm on the following page implements this theorem.

5. GENERALIZATIONS

In all of the processes discussed here except the exponential GNEAR(1) the correlations are non-negative and geometrically decreasing. A particular scheme for obtaining negative correlation is given for the exponential case. Another scheme for obtaining alternating correlations which are possibly negative and which is broadly applicable is given in Gaver and Lewis [1] and in Lawrance and Lewis [5]. Another problem is that different types of dependence and higher-order Markovian dependence might be encountered in data. Schemes for obtaining mixed autoregressive moving average exponential sequences where the autoregression has order p and the moving average has order q are given in Lawrance and Lewis [4]. The mixed exponential process TMEAR(1) is easily extended to give a process with this type of extended correlation structure. This will be discussed elsewhere.

6. ACKNOWLEDGMENTS

The research of P. A. W. Lewis was supported by the Office of Naval Research under Grant NR-42-284 at the Naval Postgraduate School.

ALGORITHM TMEAR1

This algorithm generates a sample of size N from the TMEAR1 process. It is based on the generation of arrays of uniform and unit exponential numbers. Subroutines UNIFORM and EXPON are assumed to exist to generate such arrays. For each TMEAR1 number generated two raw uniform random numbers and one exponential number are needed.

```
INPUT  N, PI1, PI2, MU1, MU2, ALPHA

MU ← PI1 * MU2 + PI2 * MU2                          /* Initialization */
 B ← MU1 + MU2 - ALPHA * MU
 A ← (1 - ALPHA) * MU1 * MU2
 T ← SQRT(B * B - 4 * A)
G2 ← .5 * (B - T)
G1 ← .5 * (B + T)
G0 ← PI2 * MU1 + PI1 * MU2
E1 ← (G1 - G0)/G1 - G2)
R1 = 1/G1
R2 = 1/G2                                           /* Initialization */
CALL UNIFORM(U,2*N+1)              /* Generate 2N+1 uniforms */
CALL EXPON(E,N+1)                  /* Generate N+1 unit exponentials */
IF  U(2N+1) ≤ PI1          THEN    X(0) ← MU1 * E(N+1)
                           ELSE    X(0) ← MU2 * E(N+1)
END IF
DO   I ← 1  to  N
     IF  U(I) ≤ E1         THEN    Z ← R1 *E(I)
                           ELSE    Z ← R2 *E(I)
     END IF
     IF  U(I+I) ≤ ALPHA    THEN    Y ← (X(I-1)
                           ELSE    Y ← 0
     END IF
     X(I) ← Y + Z
END DO
```

BIBLIOGRAPHY

[1] Gaver, D. P. and Lewis, P.A.W., "First Order Autoregressive Gamma Sequences and Point Processes," Adv. Appl. Prob. 12, 1980, 727-745.

[2] Lawrance, A. J. and Lewis, P.A.W., "A Moving Average Exponential Point Process (EMA1)," J. Appl. Prob. 14, 1977, 98-113.

[3] Jacobs, P. A. and Lewis, P.A.W., "A Mixed Autoregressive-Moving Average Exponential Sequence and Point Process (EARMA 1,1)," Adv. Appl. Prob. 9, 1977, 87-104.

[4] Lawrance, A. J. and Lewis, P.A.W., "The Exponential Autoregressive Moving Average EARMA(p,q) Process," J. R. Statist. Soc. B, 42 (2), 1980, 150-161.

[5] Lawrance, A. J. and Lewis, P.A.W., "A New Autoregressive Time-Series Model in Exponential Variables (NEAR(1))," Adv. Appl. Prob. 13 (1981) to appear.

[6] Lawrance, A.J., "The Mixed Exponential Solution to the First-Order Autoregressive Model," J. Appl. Prob. 17 1980, 546-552.

[7] Lawrance, A.J., "Some Autoregressive Models for Point Processes," Point Proccess and Queuing Problems, Colloquia Mathematica Societates Janos Bolyai, 24, P. Bartfai and J. Tomko eds., North Holland, 1981, Amsterdam, 000-000.

[8] Lewis, P.A.W., "Simple Models for Positive-Valued and Discrete-Valued Time Series with ARMA Correlation Structure," Proc. Fifth International Symposium Mult. Anal., Multivanate Analysis-V, P.R. Krishnaiah, ed. North-Holland, Amsterdam, 1980, 151-166.

[9] Lawrance, A.J. and Lewis, P.A.W., "Simulation of some Autoregressive sequences of positive random variables. Proc. 1979 Winter Simulation Conference, Highland, H.J., Spiegel, M.G. and Shannon, R. eds., 1979. IEEE Press, N.Y., 301-308.

[10] Cox, D.R. and Lewis, P.A.W., Statistical Analysis of Series of Events, Methuen, London, 1966.

[11] Lewis, P.A.W. and Shedler, G.S., "Analysis and Modelling of Point Processes in Computer Systems," Bull. ISI, XLVII (2), 1978, 193-219.

[12] Jacobs, P.A., "A Closed Cyclic Queueing Network with Dependent Exponential Service Times," J. Appl. Prob. 15, 1978, 573-589.

[13] Jacobs, P.A., "Heavy traffic results for single-server queues with dependent (EARMA) service and interarrival times," Adv. Appl. Prob. 12, 1980, 517-529.

[14] Jacobs, P.A. and Lewis, P.A.W., "Discrete Time Series Generated by Mixtures, I: Correlational and Runs Properties," J.R. Statist. Soc. B, 40(1), 1978, 94-105.

[15] Jacobs, P.A. and Lewis, P.A.W., "Discrete Time Series Generated by Mixtures II: Asymptotic Properties," J.R. Statist. Soc. B, 40(2), 1978, 222-228.

[16] Vervaat, W., "On a stochastic difference equation and a representation of non-negative infinitely divisible random variables," Adv. Appl. Prob., 11, 1979, 750-783.

[17] G. S. Fishman, personal communication.

[18] Lawrance, A.J. and Lewis, P.A.W., "A mixed exponential time series model, NMEAR(p,q)," Naval Postgraduate School Technical Report NPS55-80-012, 1980 . Submitted for publication.

DISTRIBUTION LIST

	NO. OF COPIES
Defense Technical Information Center Cameron Station Alexandria, VA 22314	2
Library, Code 0142 Naval Postgraduate School Monterey, CA 93940	2
Dean of Research Code 012 Naval Postgraduate School Monterey, CA 93940	1
Library, Code 55 Naval Postgraduate School Monterey, CA 93940	1
Professor P.A.W. Lewis Code 55Lw Naval Postgraduate School Monterey, CA 93940	274